

**FINAL OFFICE ACTION IS INAPPROPRIATE IN VIEW OF NEWLY CITED ART  
RAMAN AND KING ET AL.**

Applicants have studied the Office Action dated December 7, 2004. Applicants respectfully request entry of these remarks under the provisions of 37 C.F.R. § 1.116(a) in that the remarks below place the application and claims in condition for allowance, which allowance is respectfully requested. Claims 1 - 20 are pending. Reconsideration and allowance of the claims in view of the following remarks are respectfully requested.

As an initial matter, the Examiner made the Office Action final based on a new ground of rejection not stated in the earlier Office Action. Applicants respectfully traverse this decision. In the Final Office Action, the Examiner rejects the present claims by citing Raman (US 6,249,794), in view of King et al. (US 6,161,114), and then in further view of Meyerzon (US 6,199,081) and Meyerzon (US 6,638,314). The Applicants respectfully point out that both the Raman reference and the King reference were not cited in any of the previous Office Actions.

According to MPEP § 706.07(a): "Under present practice, second or any subsequent actions on the merits shall be final, except where the examiner introduces a new ground of rejection not necessitated by amendment of the application by applicant, whether or not the prior art is already of record." In the previous Office Action dated February 23, 2004, the Examiner rejected claims 1-3, 10, 14-16, and 20 under 35 U.S.C. §103(a) as being unpatentable over Sanu et al. (US 6,145,003), in view of Adar et al. (US 6,493,702). Also in this previous Office Action, the Examiner rejected claims 4-6, and 17-19 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (US 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Aganovic et al.

ARC9-2000-0046-US1

9

09/607,370

(US Patent Number 6,105,042). Claim 7 was rejected under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (US 6,145,003), in view of Adar et al. (US 6,493,702), in further view of Meyerzon et al. (US 6,199,081) and in further view of Meyerzon et al. (US 6,638,314). Claims 8-9 were rejected under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (US 6,493,702), and in further view of Meyerzon et al. (US 6,199,081). Claim 11 was rejected under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (US 6,145,003), in view of Adar et al. (US 6,493,702), and in further view of Meyerzon et al. (US 6,638,314). Claims 12 and 13 were rejected under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (US 6,145,003), in view of Adar et al. (US 6,493,702), in further view of Meyerzon et al. (US 6,199,314), in further view of Hughes et al. (US 5,892,908), and in further view of Aganovic (US 6,105,042).

In the previously-filed amendment, Applicants amended the claims 1 - 7, 10, and 14 - 20 for clarity. The Applicants did not switch from one subject matter to another or resort to any subterfuge to keep the application pending.<sup>1</sup> Thus it is respectfully submitted that the final status of the Office Action is premature and should be withdrawn.

If the Examiner does not withdraw the final status of the Office Action, Applicants submit that this response does not raise new issues in the application. It is submitted that the present response places the application in condition for allowance or, at least, presents the application in better form for appeal. Entry of the present response is therefore respectfully requested.

---

<sup>1</sup> See MPEP § 706.07.

### REMARKS

Applicants have studied the Office Action dated December 7, 2004. No new matter has been added. It is submitted that the application, is in condition for allowance. Applicants have amended Claims 16 and 20 for grammatical reasons. By virtue of this amendment, claims 1-20 are pending. Reconsideration and further examination of the pending claims in view of the above amendments and the following remarks is respectfully requested. In the Office Action, the Examiner:

- Rejected Claims 1-6 and 10-20 under 35 U.S.C. §103(a) as being unpatentable over Raman (U.S. Patent No. 6,249,794), in view of King et al. (U.S. Patent No. 6,161,114); and
- Rejected Claims 7-9 under 35 U.S.C. 103(a) as being unpatentable over Raman (U.S. Patent No. 6,249,794), in view of King et al. (U.S. Patent No. 6,161,114), in further view of Meyerzon et al. (U.S. Patent No. 6,199,081) and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314).

#### Telephonic Interview

Applicants wish to thank Examiner Burge and her supervisor Examiner Paula for the telephonic interview on Thursday January 27, 2004. Discussed were the technical differences of the present invention and the document description files (DTF) and how they differ from the present invention of *"loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein"*.

ARC9-2000-0046-US1

11

09/607,370

### Overview of the Present Invention

The present invention provides a web crawling unit, a method, and a computer readable medium for assisting in the index of informational content at a given website by use of a search engine. One type of search program for indexing informational content at a website is known as a "web crawler." The web crawler creates an index of informational content for a given website for subsequent use by a search engine. Simple web pages where all the content is from one or two files the indexing by web crawlers is very useful. One problem often encountered by web crawlers is where information presented in a web page is stored in secondary pages. The use of secondary documents such as text, images, and other multimedia, make the management of web page content much easier because each component of a web page is broken down into pieces. Web crawlers do not load these secondary documents when indexing web page content and this informational content is not properly indexed for a search engine.

Another problem encountered with using web crawlers to index informational content is where the web page is dynamically assembled. Dynamically assembled information is information presented on a single web page from more than one location. The dynamically assembled web content is not available from one location and must be gathered from several different storage locations before being presented as a single webpage to the user. This is especially true in websites where part of the informational content is retrieved from a database. The contents of the database must be gathered and assembled into a webpage format. Further, the use of client side scripts such as JavaScript and VBScript, do not make all the information available until the script is executed on the client side. Since web crawlers index informational content on server sites as opposed to client sites, the information content in client side script is not captured. The present invention solves the problem of web crawling dynamic data

documents by temporarily rendering the information for the dynamic data document in memory in the manner the composer of the document intended it to be displayed. In this manner, the subsequent crawling indexes all the information on the website in the manner in which the creator of the web site intended the site to be viewed. This is in contrast to only discrete pieces of the informational content on the web site being indexed in the prior art. The present invention solves the problem with images containing textual data in part through use of optical character recognition on the images themselves to assist with web crawling.

To overcome the problems of using web crawlers to properly index dynamic websites containing secondary documents, with or without client-side scripts and the use of images with textual content, the present invention retrieves a web document at a given address or URL. The contents of the web document are extracted for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Next, secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation are loaded. The in-memory webpage representation is analyzed to produce a text map for the web page document of the textual contents therein. The secondary documents include one or more images with textual content embedded therein. An optical character recognition engine is used on the images to extract textual content for adding to the textual map for the webpage document.

Rejection under 35 U.S.C. §103(a) in view of Raman and King et al.

As noted above, the Examiner rejected claims 1-6 and 10-20 under 35 U.S.C. §103(a) as being unpatentable over Raman. (U.S. Patent No. 6,249,794), in view of King et al. (U.S. Patent No. 6,161,114).

Raman discloses document description files ("DDFs") that are used to encapsulate structural and meta information associated with a document stored on a computer readable medium. See Abstract. The Examiner directs the Applicants to col. 1, lines 58-64, wherein Raman is describing a method for generating DDFs. A first DDF is generated and describes a document stored on a computer readable medium. Descriptions of the application which produced the document, the location from which the document can be obtained, and an operation which can be performed on the document to produce a second document are all included in the first DDF. The operation may extract information from the document and the second DDF may describe the extracted information or may even describe the first DDF.

The Examiner also directed the Applicants to col. 7, lines 38-51 of Raman, wherein Raman teaches how a user obtains a DDF file so that a transformation method may be applied to the DDF file. Raman teaches that a user can browse an on-line gallery using a DDF-enabled web browser, selecting a file and then selecting "Save to DDF" from the browser's menu. The DDF file is then stored on a local disk.

In contrast, as recited for independent Claim 1 and similarly for independent Claims 14 and 20, the presently claimed invention recites retrieving a web document at an

address. The contents of the web document are extracted for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit. The intermediate dynamically constructed in-memory webpage is formatted as if displayed for viewing on an end-user's web browser. Secondary documents associated with the web document are loaded in order to render the secondary documents as part of the in-memory webpage representation. The secondary documents include one or more images with textual content embedded therein. The in-memory webpage representation is analyzed and summarized to produce a text map for the web page document of the textual contents. Optical character recognition is used on the images to extract textual content for adding to the textual map for the webpage document.

Raman does not teach or suggest retrieving a web document at an address. Raman is directed towards DDFs that are used to describe the structure and content of application files. See for example, col. 4, lines 9-62. Additionally, Raman teaches how these DDFs can be used to transform files written by one application to the format used by another application. See for example, col. 7–col. 8, lines 38-67 and 1-5 respectively. Nowhere does Raman teach or suggest rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit. The invention, as taught by Raman, is not directed toward web pages and is completely absent a teaching of rendering an intermediate dynamically constructed in-memory webpage representation of the web document.

Furthermore, Raman does not teach or suggest secondary documents that are associated with the web document. As discussed above, Raman teaches DDFs, which provide an application independent description of a document saved in a native file

format of an authoring program and not web documents. See for example, col. 4, lines 25-27. Raman also teaches that a second DDF can describe a file that has been transformed from a first file format to a second file format or describe the information extracted from a source document. See for example, col. 1, lines 35-66. Alternatively, the second DDF can describe the first DDF. See for example, col. 1, lines 61-62. Nowhere does Raman teach or suggest that the secondary documents are loaded in order to render the secondary documents as part of the in-memory webpage representation.

Additionally, Raman is absent a teaching showing that an in-memory webpage representation is analyzed and summarized to produce a text map for the web page document of the textual contents and that optical character recognition is used on the images to extract textual content for adding to the textual map for the webpage document. Raman is not directed toward in-memory web-page representation and fails to teach or suggest any type of optical character recognition. The Examiner correctly states on page 5 of the present Office Action with respect to Claim 12 that Raman does not specifically mention using an optical character recognition engine. Therefore, Raman does not teach, anticipate, or suggest the presently claimed invention as recited for independent Claim 1 and similarly for independent Claims 14 and 20, and for all dependent claims depending therefrom, respectively.

King teaches a method of fitting content elements of a composition to a media layout. A document is separated into its content, design, and media aspects. King teaches automatic integration, composition, and layout of content from multiple sources into intelligent dynamic document templates that are instantly publishable in media such as print, Intranet, Internet, and in an OLE embedding. See for example, col. 2, lines 52-62.



King does not teach retrieving a web document at an address, and extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser; loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein; analyzing and summarizing the in-memory webpage representation to produce a text map for the web page document of the textual contents; and using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document, as recited for independent Claim 1 and similarly for independent Claims 14 and 20.

As stated above, King describes a process of separating document content from structure and how this can be used to lay out content on multiple output media. See for example, col. 2, lines 52-62. With respect to King, the Examiner directs the Applicants to col. 7, lines 55-65, wherein King teaches that the composition may be rendered to live HTML and merely mentions that the live HTML might incorporate JAVA applets, Shockwave objects, etc. Claim 4, on the other hand, recites "[s]econdary documents including one or more Java applets with textual content embedded therein". The secondary documents are associated with the web document and are loaded in order to render the secondary documents as part of the in-memory webpage representation. King is absent a teaching showing the above.

The Examiner further directs the Applicants to col. 12, lines 9-26 of King, wherein King teaches that content and other presentation features must be coordinated across a hierarchy of scales from individual text characters and image pixels up to the entire document, or even a series or family of documents. King also teaches that separation of content from media allows reuse of the content in different media while preserving relationships. Nowhere here nor anywhere else in King, does King teach or even suggest using optical character recognition. Accordingly, independent claims 1, 14, and 20 of the present invention distinguish over both the Raman and King references for at least the reasons stated above.

Additionally, King teaches a method of fitting content elements of a composition to a media layout. A document is separated into its content, design, and media aspects. King also teaches automatic integration, composition, and layout of content from multiple sources into intelligent dynamic document templates that are instantly publishable in media such as print, Intranet, Internet, and in an OLE embedding. The combination of Raman and King, as suggested by the Examiner, destroys the intent and purpose of "Raman taken alone or in view of King, which is the use of "DDFs" used for providing an application independent description of a document saved in a native file format of an authoring program. Accordingly, the present invention is distinguishable over Raman taken alone or in view of King for this reason as well.

The Examiner goes on to combine Raman with King.<sup>2</sup> The Examiner recites 35 U.S.C. §103. The Statute expressly requires that obviousness or non-obviousness be determined for the claimed subject matter "as a whole," and the key to proper determination of the differences between the prior art and the present invention is giving

---

<sup>2</sup> Applicants make no statement whether such combination is even proper.

full recognition to the invention "as a whole." The Raman reference taken alone or in view of King simply does not suggest, teach or disclose the patentably distinct limitations of:

an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser;

loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein;

analyzing and summarizing the in-memory webpage representation to produce a text map for the web page document of the textual contents; and

using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document.

Continuing further, when there is no suggestion or teaching in the prior art for a hub processing unit for "extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser"; "loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation"; or "using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document" the suggestion can not come from the Applicant's own specification. The Federal Circuit has repeatedly warned against using the Applicant's disclosure as a blueprint to

reconstruct the claimed invention out of isolated teachings of the prior art. See MPEP §2143 and *Grain Processing Corp. v. American Maize-Products*, 840 F.2d 902, 907, 5 USPQ2d 1788 1792 (Fed. Cir. 1988) and *In re Fitch*, 972 F.2d 160, 12 USPQ2d 1780, 1783-84 (Fed. Cir. 1992).

Moreover, the Federal Circuit has consistently held that when a §103 rejection is based upon a modification of a reference that destroys the intent, purpose or function of the invention disclosed in the reference, such a proposed modification is not proper and the *prima facie* case of obviousness can not be properly made. See *In re Gordon*, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984). Here the intent, purpose and function of Raman taken alone or in view of King is the use of "DDFs". The purpose of a DDF is to provide an application independent description of a document saved in a native file format of an authoring program. Also, Raman is directed towards the authoring stage of a file. As stated above Raman teaches how to generate a DDF and how these DDFs can be used to transform files written by one application to the format used by another application.

In contrast, the present direction is focused on files after they have been created. The intent and purpose of the present invention is "retrieving a web document at an address and extracting its contents so that an intermediate dynamically constructed in-memory webpage representation of the web document can be rendered at a hub processing unit". The web document that is retrieved by the present invention has already been created. Also, Raman does not teach or even suggest, among other things, extracting contents of a webpage so that an intermediate dynamically constructed in-memory webpage representation of the web document can be rendered at a hub processing unit.

For the foregoing reasons, independent claims 1, 14, and 20 distinguish over Raman taken alone or in view of King. Claims 2-6, 10-13, and 15-19 depend from claims 1 and 14 respectively, either directly or by way of an intervening claim. Since dependent claims contain all the limitations of the independent claims, claims 2-6, 10-13, and 15-19 distinguish over Raman taken alone and/or in view of King, as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Raman, King with Meyerzon and Meyerzon

As noted above, the Examiner rejected claims 7-9 under 35 U.S.C. 103(a) as being unpatentable over Raman. (U.S. Patent No. 6,249,794), in view of King et al. (U.S. Patent No. 6,161,114) as claimed in Claim 1, in further view of Meyerzon et al. (U.S. Patent No. 6,199,081) and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314). With respect to Raman and King, the above arguments regarding independent Claims 1, 14, and 20 are relevant here and will not be repeated. As the Examiner correctly states on page 7 of the Office Action, Raman is silent on scheduling a URL for crawling and goes on to combine Raman with Meyerzon ('081) and Meyerzon ('314).<sup>3</sup>

Additionally, the Examiner directs the Applicants to col. 8, lines 5-67 of Raman, wherein Raman teaches that the request to transform a document takes the form of an Apply-transformation element in the request DDF file. One of the two ways that the source DDF may be incorporated in the Source-DDF element is by incorporation by reference. Incorporation by reference involves placing a URL, which points to the source DDF,

---

<sup>3</sup> Applicants make no statement whether such combination is even proper.

within the Source-DDF element. Neither here nor anywhere else does Raman teach or suggest, among other things, initializing a first list with seed values; checking if there are any URLs to be processed and in response that any URL exists to be processed then performing the following sub-steps of: determining if a URL is in a second list".

Meyerzon ('081) and Meyerzon ('314) disclose a web crawler and worker thread which examines a set of properties and text within the web document. See Meyerzon ('081) at col. 9 lines 60 through col. 10, lines 27 and Meyerzon ('314) at col. 9, line 28 through col. 10, lines 12. The web crawler, as taught by Meyerzon, is not working on an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser but rather the source of the web page content itself. Accordingly, independent Claim 1 and similarly for independent Claims 14 and 20 of the present invention distinguish over Raman and/or King and/or in further view of for Meyerzon ('081) and Meyerzon ('314) for at least this reason.

For the foregoing reasons, independent Claims 1, 14, and 20 distinguish over Raman and/or King and/or in further view of for Meyerzon ('081) and Meyerzon ('314). Claims 7-9 depend from Claim 1 either directly or by way of an intervening claim. Since dependent claims contain all the limitations of the independent claims, Claims 7-9 distinguish over Raman and/or King and/or in further view of for Meyerzon ('081) and Meyerzon ('314) as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

### CONCLUSIONS

The remaining cited references have been reviewed and are not believed to effect the patentability of the claims as previously amended.

In light of the Office Action, Applicants believe these amendments serve a useful clarification purpose, and are desirable for clarification purposes, independent of patentability. Accordingly, Applicants respectfully submit that the claim amendments do not limit the range of any permissible equivalents.

Applicants acknowledge the continuing duty of candor and good faith to the disclosure of information known to be material to the examination of this application. In accordance with 37 CFR §§ 1.56, all such information is dutifully made of record. The foreseeable equivalents of any territory surrendered by amendment is limited to the territory taught by the information of record. No other territory afforded by the doctrine of equivalents is knowingly surrendered and everything else is unforeseeable at the time of this amendment by the Applicants and their attorneys.

Applicants respectfully submit that all of the grounds for rejection stated in the Examiner's Office Action have been overcome, and that all claims in the application are allowable. No new matter has been added. It is believed that the application is now in condition for allowance, which allowance is respectfully requested.

**PLEASE**, if for any reason the Examiner finds the application other than in condition for allowance, the Examiner is invited to call either of the undersigned attorneys at (561) 989-9811 should the Examiner believe a telephone interview would advance the

ARC9-2000-0046-US1


23

09/607,370

prosecution of the application.

Respectfully submitted,

Date: February 1, 2005

By:   
Jon Gibbons  
Registration No. 37,333  
Attorney for Applicants

FLEIT, KAIN, GIBBONS,  
GUTMAN, BONGINI & BIANCO P.L.  
One Boca Commerce Center, Suite 111  
551 Northwest 77th Street  
Boca Raton, FL 33487  
Tel. (561) 989-9811  
Fax (561) 989-9812

Please Direct All Future Correspondence to Customer Number **23334**

ARC9-2000-0046-US1

24

09/607,370